Quadratic minimisation problems in statistics

Albers, C. J.; Critchley, F.; Gower, J. C.

Published in:
Journal of Multivariate Analysis

DOI:
10.1016/j.jmva.2009.12.018

Publication date:
2011

Journal of Multivariate Analysis 102 (2011) 698–713

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Quadratic minimisation problems in statistics

C.J. Albers [a,b,*], F. Critchley [b], J.C. Gower [b]

[a] Psychometrics and Statistical Methods, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

[b] Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

ARTICLE INFO

Article history:
Received 15 November 2007
Available online 21 January 2010

AMS subject classifications:
15A63
15A21
62H12

Keywords:
Canonical analysis
Constraints
Geometry
Minimisation
Quadratic forms
Ratios
Reduced rank

ABSTRACT

We consider the problem $\min_{\mathbf{x}}(\mathbf{x}-\mathbf{t})'\mathbf{A}(\mathbf{x}-\mathbf{t})$ subject to $\mathbf{x}'\mathbf{B}\mathbf{x}+2\mathbf{b}'\mathbf{x}=k$ where $\mathbf{A}$ is positive definite or positive semi-definite. Variants of this problem are discussed within the framework of a general unifying methodology. These include non-trivial considerations that arise when (i) $\mathbf{A}$ and/or $\mathbf{B}$ are not of full rank and (ii) $\mathbf{t}$ takes special forms (especially $\mathbf{t}=0$ which, under further conditions, reduces to the well-known two-sided eigenvalue solution). Special emphasis is placed on insights provided by geometrical interpretations.

1. Introduction

Our main objectives in writing this paper are: (1) to bring to a statistical readership what may be some unfamiliar results to be found in a dispersed literature, (2) to unify this area, so leading to a general purpose algorithm of wide applicability, and (3) to provide some new results of our own.

The problem

$$\left.\begin{array}{l}\min\limits_{\mathbf{x}}(\mathbf{x}-\mathbf{t})'\mathbf{A}(\mathbf{x}-\mathbf{t})\\ \text{subject to } \mathbf{x}'\mathbf{B}\mathbf{x}+2\mathbf{b}'\mathbf{x}=k\end{array}\right\} \tag{1}$$

is common in statistics. In this paper we give an overview of the theory with explicit formulae that allow its immediate implementation; an exhaustive treatment with formal proofs is given by Albers et al. [1], while examples are discussed by Albers et al. [2].

We term the first part of (1) the *objective function* and the second part *the constraint*. It might arise as a problem in itself or as part of a bigger problem as, for example, when iteratively minimising a general convex objective function via a series of local quadratic approximations. Again, (1) may arise either as a direct optimisation problem with strong constraints, or as a Lagrangian form with weak constraints derived from optimising a ratio (see Gower [7] and Section 2.4 for a discussion). In the constraint, the equality can be relaxed to inequality (for exact solutions see Section 4), while additional linear equality or inequality constraints are easily subsumed. To provide ad hoc solutions to each of many special cases is inefficient. Thus, a

* Corresponding author at: Psychometrics and Statistical Methods, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.
E-mail address: c.j.albers@rug.nl (C.J. Albers).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2009.12.018

solution to (1) supporting a reliable accessible algorithm would provide a useful supplement to the toolkit of existing linear algebra algorithms.

Without loss of generality, **A** and **B** of order p, may be assumed to be symmetric. The linearly constrained case being well known, there is no real loss in also assuming that **B** is non-zero. The only restriction on the forms we allow is that **A** must be positive definite (p.d.) or positive semi-definite (p.s.d.); in this paper p.s.d. does not subsume p.d. This condition renders the objective function convex, and identifies (1) as a constrained least-squares problem. **B** is unrestricted while, often, **b** is zero, but without loss we may assume that **B** is not negative definite or negative semi-definite, if necessary by changing the signs of **B**, **b**, and k. The methodology developed below is for general k but when $k = 0$ special considerations may apply, as discussed in Section 3.3.

The minimisation problem (1) may be reparameterised in several ways. For example, it is trivial that (1) includes

$$\min_{\mathbf{x}}(\mathbf{x}-\mathbf{t})'\mathbf{A}(\mathbf{x}-\mathbf{t}) \quad \text{subject to } (\mathbf{x}-\mathbf{s})'\mathbf{B}(\mathbf{x}-\mathbf{s}) + 2\mathbf{b}'(\mathbf{x}-\mathbf{s}) = k$$

merely by setting $\mathbf{x}^* = \mathbf{x}-\mathbf{s}$ and $\mathbf{t}^* = \mathbf{t}-\mathbf{s}$. Less obvious is that the quadratically constrained full-rank regression problem below is also included:

$$\min_{\mathbf{x}} ||\mathbf{X}\mathbf{x}-\mathbf{y}||^2 \quad \text{subject to } \mathbf{x}'\mathbf{B}\mathbf{x} + 2\mathbf{b}'\mathbf{x} = k$$

but on expanding $||\mathbf{X}\mathbf{x}-\mathbf{y}||^2 = \mathbf{x}'\mathbf{X}'\mathbf{X}\mathbf{x} - 2\mathbf{y}'\mathbf{X}\mathbf{x} + \mathbf{y}'\mathbf{y}$ and then replacing $\mathbf{X}'\mathbf{X}$ by any decomposition $\mathbf{L}'\mathbf{L}$ where **L** is non-singular (e.g. use the Cholesky decomposition), we may write:

$$||\mathbf{X}\mathbf{x}-\mathbf{y}||^2 = \left(\mathbf{L}\mathbf{x} - \left(\mathbf{L}^{-1}\right)'\mathbf{X}'\mathbf{y}\right)'\left(\mathbf{L}\mathbf{x} - \left(\mathbf{L}^{-1}\right)'\mathbf{X}'\mathbf{y}\right) + \text{constant}$$

which, on defining $\mathbf{x}^* = \mathbf{L}\mathbf{x}$, $\mathbf{t} = (\mathbf{L}^{-1})'\mathbf{X}'\mathbf{y}$ and $\mathbf{A} = \mathbf{I}$, gives the basic form (1), parallel changes being made to the constraint.

We may also consider a multidimensional form of (1):

$$\left.\begin{array}{l}\min\limits_{\mathbf{X}} \text{ trace } (\mathbf{X}-\mathbf{T})'\mathbf{A}(\mathbf{X}-\mathbf{T}) \\ \text{subject to trace } (\mathbf{X}'\mathbf{B}\mathbf{X} + 2\mathbf{G}'\mathbf{X}) = k\end{array}\right\}$$

where **X**, **T** and **G** have dimensions $p \times n$. By writing $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_n')$ and similarly for **T** and **G**, and defining a block-diagonal matrix $\mathbf{A}^* = \text{diag}(\mathbf{A}, \mathbf{A}, \ldots, \mathbf{A})$ with **A** repeated n times, and similarly for $\mathbf{B}^*$, the problem becomes:

$$\left.\begin{array}{l}\min\limits_{\mathbf{x}}(\mathbf{x}-\mathbf{t})'\mathbf{A}^*(\mathbf{x}-\mathbf{t}) \\ \text{subject to } \mathbf{x}'\mathbf{B}^*\mathbf{x} + 2\mathbf{g}'\mathbf{x} = k\end{array}\right\}$$

and so is subsumed in (1).

Several special cases of (1) merit consideration: (i) $\mathbf{t} = 0$, (ii) $\mathbf{b} = 0$ and, more especially, (iii) the intersection of (i), (ii) and $k = 1$. Case (iii) is related to optimising a ratio of quadratic forms:

$$\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}}$$

as in canonical analysis, leading to the well-known two-sided eigenvalue solution, at least when one of the matrices is of full rank, but here also subsuming the case where both matrices are deficient in rank and when **B** is indefinite: see Section 2.3.

Thus, (1) includes a very wide class of problems that may manifest themselves in various equivalent forms. Some classical problems involve the simultaneous diagonalisation of **A** and **B**, at least when **A** is of full rank. Perhaps for this reason, much of the mathematical literature is concerned with conditions for simultaneous diagonalisation, but this is not always possible and is unnecessary for making progress (Section 2).

In this paper, Section 2 develops a potentially non-diagonal canonical form for **A** and **B**, also possibly including linear terms, which we term the *General Canonical Form* (GCF) and examines some important special cases of the GCF. It turns out that minimisation of all these special cases requires the solution to an analogue of a characteristic equation, the *Fundamental Canonical Equation* (FCE), whose solution is discussed in Section 3. The FCE has a unique solution lying in a simply defined admissible region or, in some special cases, at one of the extremes of this admissible region. Section 3 outlines how these results may be used to develop a general algorithm. Geometrical insights discussed in Section 4 illuminate and motivate the detailed algebraic discussion of Sections 2 and 3. Section 5 gives a discussion of related problems and extensions, some of which require further work. Proofs of some of the more technical details are given in Albers et al. [1].

2. The general canonical form

The solution to (1) is greatly simplified by recasting the problem in terms of the GCF. This depends on a simple affine transformation:

$$\mathbf{z} = \mathbf{T}^{-1}\mathbf{x} + \mathbf{m} \tag{2}$$

where **T** and **m** are chosen to simultaneously simplify $\mathbf{T}'\mathbf{A}\mathbf{T}$ and $\mathbf{T}'\mathbf{B}\mathbf{T}$; in some, but not all, cases this becomes simultaneous diagonalisation.

We use the inverse $\mathbf{T}^{-1}$ here because it turns out to simplify the appearance of $\mathbf{T}$. Explicit forms of the non-singular matrix $\mathbf{T}$ and the translation $\mathbf{m}$ are given in (5) and the discussion leading to (9). First, we need to develop some notation.

The transformation $\mathbf{T}$ may be written explicitly in terms of three spectral decompositions shown in (3), the decomposition of $\mathbf{A}$ being used to define $\mathbf{C}$ in (4) and, thereby, the other two decompositions. Denoting non-null eigenvectors by a unit suffix (e.g. $\mathbf{U}_1$) and null eigenvectors by a zero suffix (e.g. $\mathbf{U}_0$) with $\mathbf{U} = \begin{pmatrix}\mathbf{U}_1 & \mathbf{U}_0\end{pmatrix}$ giving a complete orthogonal matrix, the required spectral decompositions are as follows:

$$\left.\begin{array}{l}\mathbf{A} = \mathbf{U}_1\Delta^2\mathbf{U}_1' \\ \mathbf{C}_{22} = \mathbf{W}_1\Gamma_0\mathbf{W}_1' \\ \mathbf{C}_{11} - \mathbf{C}_{12}\left(\mathbf{W}_1\Gamma_0^{-1}\mathbf{W}_1'\right)\mathbf{C}_{21} = \mathbf{V}_1\Gamma_1\mathbf{V}_1'\end{array}\right\} \tag{3}$$

where

$$\mathbf{C} = \begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{pmatrix} = \begin{pmatrix}\Delta^{-1}\mathbf{U}_1'\mathbf{B}\mathbf{U}_1\Delta^{-1} & \Delta^{-1}\mathbf{U}_1'\mathbf{B}\mathbf{U}_0 \\ \mathbf{U}_0'\mathbf{B}\mathbf{U}_1\Delta^{-1} & \mathbf{U}_0'\mathbf{B}\mathbf{U}_0\end{pmatrix}. \tag{4}$$

$\Delta^2, \Gamma_1, \Gamma_0$ are all non-singular diagonal matrices of eigenvalues; we have used Δ^2 to emphasise the positivity of the non-zero eigenvalues of $\mathbf{A}$. Also, define $\mathbf{D}_{10} = \mathbf{V}_1'\mathbf{C}_{12}\mathbf{W}_0$ and $\mathbf{D}_{00} = \mathbf{V}_0'\mathbf{C}_{12}\mathbf{W}_0$.

With the notation given in (3) and (4), the transformation may be written explicitly as:

$$\mathbf{T} = \begin{pmatrix}\mathbf{U}_1\Delta^{-1}\mathbf{V} - \mathbf{U}_0\mathbf{W}_1\Gamma_0^{-1}\mathbf{W}_1\mathbf{C}_{21}\mathbf{V} & \mathbf{U}_0\mathbf{W}\end{pmatrix} \tag{5}$$

with inverse

$$\mathbf{T}^{-1} = \begin{pmatrix}\mathbf{V}'\Delta\mathbf{U}_1' \\ \mathbf{W}'\mathbf{W}_1\Gamma_0^{-1}\mathbf{W}_1'\mathbf{C}_{21}\Delta\mathbf{U}_1' + \mathbf{W}'\mathbf{U}_0'\end{pmatrix}. \tag{6}$$

Main result

There exists a non-singular transformation $\mathbf{T}$ that gives the General Canonical Form:

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \left(\begin{array}{cc|c} \mathbf{I} & & \\ & \mathbf{I} & \\ \hline & & \end{array}\right); \qquad \mathbf{T}'\mathbf{B}\mathbf{T} = \left(\begin{array}{cc|c} \Gamma_1 & & \mathbf{D}_{10} \\ & & \mathbf{D}_{00} \\ \hline \mathbf{D}_{10}' & \mathbf{D}_{00}' & \Gamma_0 \end{array}\right). \tag{7}$$

We believe that (5) and (6) are new. We do not give a detailed derivation of (7) here, but on substituting for $\mathbf{T}$ from (5) it is easily, if somewhat tediously, verified to give the result claimed. A constructive proof may be found in Albers et al. [1]. Note that because the null eigenvectors (those with zero suffices) are determined only up to spanning the null space, there is a degree of non-uniqueness in (5) and (6). This has no substantive effect in applications.

Underlying the GCF is the interplay of the range and null spaces of $\mathbf{A}$ and $\mathbf{B}$. The vector $\mathbf{z}$ is decomposed into components $\mathbf{z} = (\mathbf{z}_{11}', \mathbf{z}_{10}', \mathbf{z}_{01}', \mathbf{z}_{00}')$ where the first suffix position refers to $\mathbf{A}$ and the second to $\mathbf{B}$; a suffix 1 denotes the range space and 0 the null space. We term the variables $\mathbf{z}_{01}$ and $\mathbf{z}_{00}$ that occur only in the constraint, *extraneous variables*.

In (7) we have indicated a partition into the range and null spaces of $\mathbf{A}$. The implicit partitions corresponding to the diagonal matrices Γ_1 and Γ_0 are in the range space of $\mathbf{T}'\mathbf{B}\mathbf{T}$ although, as explained below, when $\mathbf{B}$ is indefinite additional parts of its range space may correspond to zero diagonal values. The range space of $\mathbf{A}$, and similar matrices, refers to the substantive dimensionality of the quadratic form $\mathbf{z}'\mathbf{A}\mathbf{z}$; for all practical purposes, this means the space spanned by the eigenvectors of $\mathbf{A}$ corresponding to non-zero eigenvalues.

For p.s.d. matrices, $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$ implies that $\mathbf{B}\mathbf{x} = 0$, or equivalently $\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{z} = 0$, so then $\mathbf{D}_{00}$ and $\mathbf{D}_{10}$ vanish, the zero diagonals correspond to the null space of $\mathbf{B}$. and Γ_1 and Γ_0 account for the whole range space of $\mathbf{B}$. However, for indefinite matrices this is not so and the partitions corresponding to the zero diagonals may include parts of the range space of $\mathbf{B}$. To be more explicit and using the notation that ζ_{ij} denotes the vector $\mathbf{z}' = (\mathbf{z}_{11}', \mathbf{z}_{10}', \mathbf{z}_{01}', \mathbf{z}_{00}')$ with all elements zero except $\mathbf{z}_{ij}$ $(i = 0, 1; j = 0, 1)$, we note that when $\mathbf{D}_{00}$ and/or $\mathbf{D}_{10}$ do not vanish, $(\mathbf{T}'\mathbf{B}\mathbf{T})\zeta_{10}$ and $(\mathbf{T}'\mathbf{B}\mathbf{T})\zeta_{00}$ need not be null even though the diagonal values $\zeta_{10}'(\mathbf{T}'\mathbf{B}\mathbf{T})\zeta_{10}$ and $\zeta_{00}'(\mathbf{T}'\mathbf{B}\mathbf{T})\zeta_{00}$ both vanish.

Linear terms and expressing (1) in terms of the GCE

The above suffices for defining all the terms in the GCF (7). We next examine how this simplifies the minimisation problem (1) and also what, if any, further simplifications can be made to the linear term in the constraint.

The transformation $\mathbf{T}$ operates on $\mathbf{x}$ and $\mathbf{t}$ to give $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x} = (\mathbf{z}_{11}', \mathbf{z}_{10}', \mathbf{z}_{01}', \mathbf{z}_{00}')'$ and $\mathbf{s} = \mathbf{T}^{-1}\mathbf{t}$. We may also write $\mathbf{b}'\mathbf{x} = \mathbf{g}'\mathbf{z}$ where $\mathbf{g}' = \mathbf{b}'\mathbf{T}$. The effect of the transformation $\mathbf{T}$ is that (1) simplifies to:

$$\left.\begin{array}{l}\min\limits_{\mathbf{z}} ||\mathbf{z}_{11} - \mathbf{s}_{11}||^2 + ||\mathbf{z}_{10} - \mathbf{s}_{10}||^2 \\ \text{subject to } \mathbf{z}_{11}'\Gamma_1\mathbf{z}_{11} + \mathbf{z}_{01}'\Gamma_0\mathbf{z}_{01} + 2\mathbf{z}_{11}'\mathbf{D}_{10}\mathbf{z}_{00} + 2\mathbf{z}_{10}'\mathbf{D}_{00}\mathbf{z}_{00} + 2\mathbf{g}'\mathbf{z} = k\end{array}\right\}. \tag{8}$$

The linear term in (8) may be simplified further. Examining the individual contributions gives $\mathbf{g}'\mathbf{z} = \mathbf{g}_{11}'\mathbf{z}_{11} + \mathbf{g}_{10}'\mathbf{z}_{10} + \mathbf{g}_{01}'\mathbf{z}_{01} + \mathbf{g}_{00}'\mathbf{z}_{00}$ and focussing on the term $\mathbf{g}_{11}'\mathbf{z}_{11}$, we observe that:

$$\mathbf{z}_{11}'\Gamma_1\mathbf{z}_{11} + 2\mathbf{g}_{11}'\mathbf{z}_{11} - k = \left(\mathbf{z}_{11} + \Gamma_1^{-1}\mathbf{g}_{11}\right)'\Gamma_1\left(\mathbf{z}_{11} + \Gamma_1^{-1}\mathbf{g}_{11}\right) - \left(k + \mathbf{g}_{11}'\Gamma_1^{-1}\mathbf{g}_{11}\right).$$

A similar observation applies to $\mathbf{g}'_{01}\mathbf{z}_{01}$, allowing the linear terms in $\mathbf{g}_{11}$ and $\mathbf{g}_{01}$ to be absorbed into the quadratic terms, by the translation

$$\mathbf{z} \to \mathbf{z} + \mathbf{m} \quad \text{where } \mathbf{m}' = \left(\mathbf{g}'_{11}\Gamma_1^{-1}, \quad \mathbf{0}, \quad \mathbf{g}'_{01}\Gamma_0^{-1}, \quad \mathbf{0}\right).$$

The same translation must be made to $\mathbf{s}_{11}$ but not $\mathbf{s}_{01}$ (which is undefined). Combining all these refinements, (8) becomes:

$$\left.\begin{array}{l} \min\limits_{\mathbf{z}} ||\mathbf{z}_{11} - \mathbf{s}_{11}||^2 + ||\mathbf{z}_{10} - \mathbf{s}_{10}||^2 \\ \text{subject to} \\ \mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{z}'_{11}\mathbf{D}_{10}\mathbf{z}_{00} + 2\mathbf{z}'_{10}\mathbf{D}_{00}\mathbf{z}_{00} + 2\mathbf{g}'_{10}\mathbf{z}_{10} + 2\mathbf{g}'_{00}\mathbf{z}_{00} = k \end{array}\right\} \tag{9}$$

where it is assumed that, when appropriate, the translational adjustments are included in $\mathbf{z}_{11}, \mathbf{z}_{01}, \mathbf{s}_{11}$ and k. In matrix form (9) is:

$$\left.\begin{array}{l} \min\limits_{\mathbf{z}} \left(||\mathbf{z}_{11} - \mathbf{s}_{11}||^2 + ||\mathbf{z}_{10} - \mathbf{s}_{10}||^2\right) \\ \text{subject to:} \\ \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{10} \\ \mathbf{z}_{01} \\ \mathbf{z}_{00} \end{pmatrix}' \begin{pmatrix} \Gamma_1 & & & \mathbf{D}_{10} \\ & \mathbf{0} & & \mathbf{D}_{00} \\ & & \Gamma_0 & \\ \mathbf{D}'_{10} & \mathbf{D}'_{00} & & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{z}_{11} \\ \mathbf{z}_{10} \\ \mathbf{z}_{01} \\ \mathbf{z}_{00} \end{pmatrix} + 2 \begin{pmatrix} \mathbf{g}_{10} \\ \\ \mathbf{g}_{00} \end{pmatrix}' \begin{pmatrix} \mathbf{z}_{10} \\ \\ \mathbf{z}_{00} \end{pmatrix} = k \end{array}\right\} \tag{10}$$

where $\mathbf{s} = \mathbf{T}^{-1}\mathbf{t} + \mathbf{m}$. Γ_1 and Γ_0 are diagonal matrices defined in (3) and (4) in the range space of $\mathbf{B}$ associated with the range and null spaces, respectively, of $\mathbf{A}$. All terms in (10) are linear functions of the parameters of (1). To minimise (1) it suffices to minimise (10), substituting into (2) to derive the optimal setting of $\mathbf{x}$.

For many problems encountered in practice, the full form of (10) simplifies. Exact solutions $\mathbf{z}_{11} = \mathbf{s}_{11}, \mathbf{z}_{10} = \mathbf{s}_{10}$ may exist, but when they do, this suggests that too loose constraints are being applied. The main special cases that occur in practice are addressed below.

Eq. (9) seems to be the simplest general form of (1) though further simplifications occur in important special cases, considered in the following. Nothing is lost by considering the matrix in (10) to be the transformed version of $\mathbf{B}$, so in the remainder of this Section references to $\mathbf{B}$ are to its transformed, or canonical, form.

2.1. Case 1. $\mathbf{A}$ *of full rank*

When $\mathbf{A}$ is of full rank, it has no null space so the extraneous variables $\mathbf{z}_{01}$ and $\mathbf{z}_{00}$ do not occur. Note that $\mathbf{B}$ is allowed to be indefinite. This is the most important practical case because it implies that, as is usually the case, the constraint contains no variables that are not in the objective function. Thus $\mathbf{z}' = (\mathbf{z}'_{11}, \mathbf{z}'_{10})$, and the minimisation problem (10) reduces to:

$$\left.\begin{array}{l} \min\limits_{\mathbf{z}} \left(||\mathbf{z}_{11} - \mathbf{s}_{11}||^2 + ||\mathbf{z}_{10} - \mathbf{s}_{10}||^2\right) \\ \text{subject to } \mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + 2\mathbf{g}'_{10}\mathbf{z}_{10} = k \end{array}\right\}.$$

The normal equations derived from the Lagrangian are:

$$\left.\begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda\Gamma_1\mathbf{z}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda\mathbf{g}_{10} \end{array}\right\}. \tag{11}$$

When there is no linear term, the second equation of (11) gives $\mathbf{z}_{10} = \mathbf{s}_{10}$ and the problem is in what we term the fundamental canonical form (FCF) discussed in Section 3 which gives the solution to the minimisation of the FCF when exact solutions are unavailable. Usually, but not always, the associated fundamental canonical equation (FCE) has a unique easily computed minimum (see Section 3). When the linear term $\mathbf{g}_{10}$ is included, the constraint becomes:

$$\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + 2\lambda\mathbf{g}'_{10}\mathbf{g}_{10} = k - 2\mathbf{g}'_{10}\mathbf{s}_{10}. \tag{12}$$

Note the positive coefficient of λ, shown in Section 3 to have little affect on solving the FCE.

2.2. Case 2. $\mathbf{A}$ *not of full rank,* $\mathbf{B}$ *is diagonal (including* $\mathbf{B}$ *positive semi-definite)*

So far we have assumed that there are no extraneous variables, $\mathbf{z}_{01}$ and $\mathbf{z}_{00}$. In Case 2 we begin to relax this condition. Suppose that $\mathbf{B}$ is p.s.d. The GCF simplifies because any zero diagonal values of a p.s.d. matrix induce corresponding entire zero rows and columns. Thus, in (10) the matrices $\mathbf{D}_{10}$ and $\mathbf{D}_{00}$ are zero. Also, both Γ_1 and Γ_0 must be positive definite. Thus, in this case (10) represents a simultaneous diagonalisation of $\mathbf{A}$ and $\mathbf{B}$ [14]. Diagonal forms of (10) may also occur when $\mathbf{B}$ is not p.s.d. in which case Γ_1 and/or Γ_0 must be indefinite. We now consider all these situations.

First, consider the case where $\mathbf{B}$ is p.s.d. There is an exact solution with $\mathbf{z}_{11} = \mathbf{s}_{11}, \mathbf{z}_{10} = \mathbf{s}_{10}$ when the constraint

$$\mathbf{s}'_{11}\Gamma_1\mathbf{s}_{11} + 2\mathbf{g}'_{10}\mathbf{s}_{10} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{g}'_{00}\mathbf{z}_{00} = k \tag{13}$$

has solutions for $\mathbf{z}_{01}$ and $\mathbf{z}_{00}$. For any given setting of $\mathbf{z}_{01}$, and when $\mathbf{g}_{00} \neq 0$, (13) always has a solution for $\mathbf{z}_{00}$. Therefore, we regard the problem as overparameterised if the linear term in $\mathbf{g}_{00}$ is admitted. Even when $\mathbf{g}_{00} = 0$ (13) may still have exact solutions but in general cannot be satisfied and we must seek non-exact solutions. Then, the normal equations of the Lagrangian simplify to:

$$\left.\begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda\Gamma_1\mathbf{z}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda\mathbf{g}_{10} \\ 0 = \lambda\Gamma_0\mathbf{z}_{01} \\ 0 = \lambda\mathbf{g}_{00} \end{array}\right\}. \tag{14}$$

When there is no exact solution, the first two equations of (14) show that we must have $\lambda \neq 0$. Then, the last equation of (14) shows that $\mathbf{g}_{00} = 0$, confirming what we have already seen above that $\mathbf{g}_{00} = 0$ is a necessary condition for a non-exact solution. When present, the third equation of (14) gives $\mathbf{z}_{01} = 0$. Thus, the extraneous variables are either absent, zero ($\mathbf{z}_{01}$) or irrelevant ($\mathbf{z}_{00}$). Overall, we arrive back at the Eq. (11) of the FCF.

When $\mathbf{B}$ is diagonal but not p.s.d. the only difference is that (13) has exact solutions whenever Γ_0 is indefinite as well as when $\mathbf{g}_{00} \neq 0$. We are therefore led to regard this situation to be overparameterised and admit only Γ_1 to be indefinite and Γ_0 to be either positive or negative definite, depending on the sign of $k - \mathbf{s}'_{11}\Gamma_1\mathbf{s}_{11} - 2\mathbf{g}'_{10}\mathbf{s}_{10}$.

2.3. Case 3. **A** *not of full rank,* **B** *not diagonal*

In this case we must treat the full GCF, first examining the possibility of exact solutions. We have seen that when $\mathbf{B}$ is p.s.d., it must be diagonal. Conversely, when $\mathbf{B}$ includes non-zero $\mathbf{D}_{10}$ and/or $\mathbf{D}_{00}$ it must be indefinite. The constraint may be written:

$$\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{g}'_{10}\mathbf{z}_{10} + 2\mathbf{p}'(\mathbf{z})\mathbf{z}_{00} = k, \tag{15}$$

where $\mathbf{p}(\mathbf{z}) = \mathbf{D}'_{10}\mathbf{z}_{11} + \mathbf{D}'_{00}\mathbf{z}_{00} + \mathbf{g}_{00}$. For an exact solution, $\mathbf{z}_{11} = \mathbf{s}_{11}$ and $\mathbf{z}_{10} = \mathbf{s}_{10}$, and (15) becomes:

$$\mathbf{s}'_{11}\Gamma_1\mathbf{s}_{11} + \mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} + 2\mathbf{g}'_{10}\mathbf{s}_{10} + 2\mathbf{p}'(\mathbf{s})\mathbf{z}_{00} = k, \tag{16}$$

which must have real solutions for $\mathbf{z}_{00}$ and $\mathbf{z}_{01}$. For any setting of $\mathbf{z}_{01}$, (16) is linear in $\mathbf{z}_{00}$ and, unless $\mathbf{p}(\mathbf{s}) = 0$, will always have a solution. When $\mathbf{p}(\mathbf{s}) = 0$, $\mathbf{z}_{00}$ becomes arbitrary but there remains the possibility that $\mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01} = k - \mathbf{s}'_{11}\Gamma_1\mathbf{s}_{11} - 2\mathbf{g}'_{10}\mathbf{s}_{10}$ has solutions for $\mathbf{z}_{01}$; this is trivially true when Γ_0 is itself indefinite and remains a possibility when Γ_0 is definite. It follows that for exact solutions not to exist it is necessary, though not sufficient, for Γ_0 to be definite or absent, i.e. $\mathbf{z}_{01}$ is excluded from the constraint. Furthermore, we require $\mathbf{p}(\mathbf{s}) = 0$. With these settings, $\mathbf{z}_{00}$ enters neither into the objective function nor the constraint. Nevertheless, as we see below in (17), $\mathbf{z}_{00}$ still enters the normal equations. For a full treatment of essentially exact solutions, see Albers et al. [1].

When there is a non-exact solution, the normal equations will take the following form for non-zero λ:

$$\left.\begin{array}{l} \mathbf{z}_{11} - \mathbf{s}_{11} = \lambda(\Gamma_1\mathbf{z}_{11} + \mathbf{D}_{10}\mathbf{z}_{00}) \\ \mathbf{z}_{10} - \mathbf{s}_{10} = \lambda(\mathbf{D}_{00}\mathbf{z}_{00} + \mathbf{g}_{10}) \\ 0 = \lambda\Gamma_0\mathbf{z}_{01} \\ 0 = \lambda(\mathbf{D}'_{10}\mathbf{z}_{11} + \mathbf{D}'_{00}\mathbf{z}_{10} + \mathbf{g}_{00}) \end{array}\right\}. \tag{17}$$

The third of the Eqs. (17) shows that, when present, $\mathbf{z}_{01} = 0$ is necessary for a non-exact solution. The fourth equation gives $\mathbf{p}(\mathbf{z}) = 0$, to which must be added the necessary condition $\mathbf{p}(\mathbf{s}) = 0$ for a non-exact solution. The constraint simplifies to:

$$\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11} + 2\mathbf{z}'_{10}\mathbf{g}_{10} = k. \tag{18}$$

From $\mathbf{p}(\mathbf{z}) - \mathbf{p}(\mathbf{s}) = 0$ we have that $\mathbf{D}'_{10}(\mathbf{z}_{11} - \mathbf{s}_{11}) + \mathbf{D}'_{00}(\mathbf{z}_{10} - \mathbf{s}_{10}) = 0$ which shows that $\mathbf{z}_{11} - \mathbf{s}_{11}$ and $\mathbf{z}_{10} - \mathbf{s}_{10}$ must lie in the null row space of $\mathbf{D}_{10}$ and $\mathbf{D}_{00}$. That is

$$\begin{pmatrix} \mathbf{z}_{11} - \mathbf{s}_{11} \\ \mathbf{z}_{10} - \mathbf{s}_{10} \end{pmatrix} = \begin{pmatrix} \mathbf{H}_1\mathbf{h} \\ \mathbf{H}_0\mathbf{h} \end{pmatrix} \quad \text{where } \mathbf{H}'\mathbf{D} = \begin{pmatrix} \mathbf{H}'_1 & \mathbf{H}'_0 \end{pmatrix}\begin{pmatrix} \mathbf{D}_{10} \\ \mathbf{D}_{00} \end{pmatrix} = 0 \tag{19}$$

for some $\mathbf{h}$. The matrices $\mathbf{H}_1$ and $\mathbf{H}_0$ may be found from the SVD of $\mathbf{D}$ and without loss of generality $\mathbf{H}$ may be assumed to be column orthonormal. The first two equations of (17) then become:

$$\left.\begin{array}{l} \mathbf{H}_1\mathbf{h} = \lambda(\Gamma_1(\mathbf{H}_1\mathbf{h} + \mathbf{s}_{11}) + \mathbf{D}_{10}\mathbf{z}_{00}) \\ \mathbf{H}_0\mathbf{h} = \lambda(\mathbf{D}_{00}\mathbf{z}_{00} + \mathbf{g}_{10}) \end{array}\right\}.$$

Multiplying the first equation by $\mathbf{H}'_1$ the second by $\mathbf{H}'_0$ and adding, using (19) and orthonormality of $\mathbf{H}$, gives:

$$\mathbf{h} = \lambda\left((\mathbf{H}'_1\Gamma_1\mathbf{H}_1)\mathbf{h} + \mathbf{g}\right) \tag{20}$$

where $\mathbf{g} = \mathbf{H}'_1\Gamma_1\mathbf{s}_{11} + \mathbf{H}'_0\mathbf{g}_{10}$.

The constraint (18) becomes:

$$(\mathbf{H}_1\mathbf{h}+\mathbf{s}_{11})'\Gamma_1(\mathbf{H}_1\mathbf{h}+\mathbf{s}_{11})+2\mathbf{g}'_{10}(\mathbf{H}_0\mathbf{h}+\mathbf{s}_{10})=k,$$

i.e.

$$\mathbf{h}'(\mathbf{H}'_1\Gamma_1\mathbf{H}_1)\mathbf{h}+2\mathbf{h}'\mathbf{g}=k-\mathbf{s}'_{11}\Gamma_1\mathbf{s}_{11}-2\mathbf{g}'_{10}\mathbf{s}_{10}=k^*, \tag{21}$$

say.

With the spectral decomposition $\mathbf{H}'_1\Gamma_1\mathbf{H}_1=\mathbf{K}'\Delta\mathbf{K}$, (20) and (21) give:

$$\left.\begin{array}{l}\mathbf{Kh}=\lambda(\Delta\mathbf{Kh}+\mathbf{Kg})\\(\mathbf{Kh})'\Delta(\mathbf{Kh})+2(\mathbf{Kh})'(\mathbf{Kg})=k^*\end{array}\right\}. \tag{22}$$

With $\mathbf{z}=\mathbf{Kh}$ and $\mathbf{b}=\mathbf{Kg}$, these are the normal equations arising from minimisation of $\|\mathbf{z}\|^2$ subject to $\mathbf{z}'\Delta\mathbf{z}+2\mathbf{b}'\mathbf{z}=k^*$. Since Δ is diagonal, this is an instance of Case 2 above. Its solution $\hat{\mathbf{z}}$ gives $\hat{\mathbf{h}}=\mathbf{K}'\hat{\mathbf{z}}$ which may be substituted into (19) to obtain $\hat{\mathbf{z}}_{11}$ and $\hat{\mathbf{z}}_{10}$.

In general, the constraint $\mathbf{p}(\mathbf{s})=0$ for non-exact solution seems unrealistic. An exception is when $\mathbf{D}_{10}$ and $\mathbf{D}_{00}$ are both zero, in which case we return to Case 2.

2.4. Ratios of quadratic forms

In this section we shall assume that the matrices $\mathbf{A}$ and $\mathbf{B}$ have already been transformed into their GCF. Often we require to minimise the ratio:

$$\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Az}} \tag{23}$$

which in its most general form becomes

$$\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Az}}=\frac{\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11}+\mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01}+2\mathbf{z}'_{11}\mathbf{D}_{11}\mathbf{z}_{00}+2\mathbf{z}'_{10}\mathbf{D}_{10}\mathbf{z}_{00}}{\mathbf{z}'_{11}\mathbf{z}_{11}+\mathbf{z}'_{10}\mathbf{z}_{10}}. \tag{24}$$

With this generality, the ratio may be made infinite by choosing $\mathbf{z}_{11}=\mathbf{0}$, $\mathbf{z}_{10}=\mathbf{0}$, and $\mathbf{z}_{01}\neq\mathbf{0}$. Further, its sign will depend on whether $\mathbf{z}_{01}$ is made to match a positive or negative element of Γ_0. The ratio may be made zero by choosing $\mathbf{z}_{11}=\mathbf{0}$, $\mathbf{z}_{00}=\mathbf{0}$, $\mathbf{z}_{01}=\mathbf{0}$, and $\mathbf{z}_{10}\neq\mathbf{0}$.

Thus, in general the ratio does not have an interesting maximum or minimum, except in special cases discussed in the following. For example, the classical special case is when $\mathbf{A}$ is p.d. and $\mathbf{B}$ p.s.d., in which case the ratio simplifies to:

$$\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Az}}=\frac{\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11}}{\mathbf{z}'_{11}\mathbf{z}_{11}}$$

with a maximum of γ_p and minimum of γ_1.

An interesting special case if when $\mathbf{A}$ and $\mathbf{B}$ are both p.s.d., giving

$$\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Az}}=\frac{\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11}+\mathbf{z}'_{01}\Gamma_0\mathbf{z}_{01}}{\mathbf{z}'_{11}\mathbf{z}_{11}+\mathbf{z}'_{00}\mathbf{z}_{00}}$$

which may be made zero (by choosing $\mathbf{z}_{11}=\mathbf{0}$, $\mathbf{z}_{01}=\mathbf{0}$ and $\mathbf{z}_{10}\neq\mathbf{0}$) or infinite ($\mathbf{z}_{11}=\mathbf{0}$, $\mathbf{z}_{10}=\mathbf{0}$ and $\mathbf{z}_{01}\neq\mathbf{0}$). This too seems uninteresting but when we may express $\mathbf{A}=\mathbf{B}+\mathbf{W}$, where all three matrices are p.s.d. we have that

$$\mathbf{z}'\mathbf{Az}=\mathbf{z}'\mathbf{Bz}+\mathbf{z}'\mathbf{Wz}$$

and it follows that when $\mathbf{z}$ is a null-vector of $\mathbf{A}$ so it must be of $\mathbf{B}$ (and $\mathbf{W}$). Thus, the null space of $\mathbf{B}$ must include the null space of $\mathbf{A}$, though it may also have additional null vectors. This implies that the rows and columns of the GCF that are associated with $\mathbf{z}_{01}$ are null, and so

$$\frac{\mathbf{z}'\mathbf{Bz}}{\mathbf{z}'\mathbf{Az}}=\frac{\mathbf{z}'_{11}\Gamma_1\mathbf{z}_{11}}{\mathbf{z}'_{11}\mathbf{z}_{11}+\mathbf{z}'_{10}\mathbf{z}_{10}}$$

which is zero for $\mathbf{z}_{11}=\mathbf{0}$ and $\mathbf{z}_{10}\neq\mathbf{0}$ but is never infinite. A maximum γ_p now occurs by choosing $\mathbf{z}_{11}$ to match Γ_1 and $\mathbf{z}_{10}=0$. This is an important special case, for the condition $\mathbf{A}=\mathbf{B}+\mathbf{W}$ is satisfied when $\mathbf{A}$ is a "total" dispersion matrix, $\mathbf{B}$ a "between group" dispersion matrix and $\mathbf{W}$ a "within group" dispersion matrix, an analysis of variance that underlies several forms of statistical canonical analysis.

When $\mathbf{B}$ is indefinite we may still have that $\mathbf{T}'\mathbf{BT}$ is diagonal with either or both of Γ_1 and Γ_0 indefinite, generally leading to zero-ratio solutions. Further, we may consider the off-diagonal matrices $\mathbf{D}_1$ and $\mathbf{D}_0$ of the GCF. We echo the comment of de Leeuw [10]: "if the pair $\mathbf{A}$ and $\mathbf{B}$ is not simultaneously diagonalisable, then the situation becomes considerably more complicated [and] the relevance of this case for practical data is limited", see also Section 2.3.

In the above, the scaling of $\mathbf{z}$ is irrelevant, but if we seek multiple solutions based on two or more eigenvalues, as is commonplace in canonical analysis, then the relative scaling becomes important. Usually, this is handled by imposing a matrix constraint of the form $\mathbf{Z}'\mathbf{AZ}=\mathbf{I}$. This differs from the scalar constraint of (1) and introduces considerations beyond the scope of this paper (see Albers et al., [2] for a discussion).

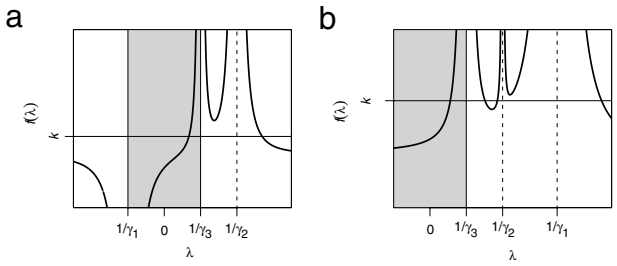

Fig. 1. (a) The form of $f(\lambda)$ when $\mathbf{B}$ is indefinite. Asymptotes occur for positive and negative values of λ. In the shaded region F, $f(\lambda)$ is monotone increasing so contains a unique root. (b) The form of $f(\lambda)$ when $\mathbf{B}$ is p.s.d. All asymptotes are for positive λ and F stretches to $-\infty$.

3. Solving the fundamental canonical equation

We have seen in Section 2 that the basic minimisation problem required by all forms of (1) may be expressed in the Fundamental Canonical Form (FCF)

$$\left.\begin{array}{l}\min\limits_{\mathbf{z}} ||\mathbf{z}-\mathbf{s}||^2 \\ \text{subject to } \mathbf{z}'\Gamma\mathbf{z}=k\end{array}\right\}, \tag{25}$$

with Γ non-singular (noting that k here differs from k in (1) unless $\mathbf{b}=0$).

The Lagrangian form is to minimise:

$$||\mathbf{z}-\mathbf{s}||^2-\lambda(\mathbf{z}'\Gamma\mathbf{z}-k)$$

which on differentiation, gives:

$$(\mathbf{z}-\mathbf{s})-\lambda\Gamma\mathbf{z}=0.$$

When $\mathbf{s}'\Gamma\mathbf{s}=k$ the constraint is satisfied for the exact solution $\mathbf{z}=\mathbf{s}$ and $\lambda=0$. For approximate solutions,

$$\mathbf{z}=(\mathbf{I}-\lambda\Gamma)^{-1}\mathbf{s},$$

which on substitution into the constraint gives:

$$\mathbf{s}'(\mathbf{I}-\lambda\Gamma)^{-1}\Gamma(\mathbf{I}-\lambda\Gamma)^{-1}\mathbf{s}=k. \tag{26}$$

Eq. (26) gives the basic Fundamental Canonical Equation (FCE) that has to be solved for λ. It is convenient to write (26) in the non-matrix form:

$$f(\lambda)=\sum_{i=1}^{p}\frac{\gamma_i s_i^2}{(1-\lambda\gamma_i)^2}=k. \tag{27}$$

The FCE could be expanded as a polynomial of degree $2p$, but it is more convenient to retain its original form. Provided $s_i\neq 0$, there are vertical asymptotes at $\lambda=1/\gamma_i$ $(i=1,2,\ldots,p)$; the case where some, or all s_i are zero, is discussed below. Assuming that the eigenvalues γ_i are given in increasing order, for indefinite $\mathbf{B}$ γ_1 will be the smallest negative eigenvalue and γ_p the largest positive eigenvalue. However, for p.d matrices $\mathbf{B}$, all eigenvalues will be positive.

Fig. 1a indicates the general shape of $f(\lambda)$ when $\mathbf{B}$ is indefinite. We see that the origin is contained in the shaded interval:

$$\mathrm{F}:\lambda\in\left(\frac{1}{\gamma_1}\leq 0\leq\frac{1}{\gamma_p}\right),$$

termed the *admissible region* for reasons about to be explained. Furthermore the figure shows that $f(\lambda)$ is strictly monotonically increasing in this interval and not in any other interval determined by adjacent asymptotes. This is an easily proved general result. Because of the monotonicity, there is at most one root in F. There may be real pairs of roots in other intervals but from the Hessian $\mathbf{H}=\mathbf{I}-\lambda\Gamma$, it follows that it is only when $\lambda\in$ F are all the diagonal values of $\mathbf{H}$ positive, indicating the existence of a minimum; other roots all refer to saddle-point solutions or, possibly, a root indicating a maximum in one of the end branches of $f(\lambda)$. Thus, it suffices to focus on the admissible region F, shaded in Fig. 1a. When $\mathbf{B}$ is p.s.d. no γ_i is negative and then Fig. 1b illustrates the behaviour of $f(\lambda)$. The only change is that F now extends to $-\infty$. The Algorithm Section in Section 3.3 discusses algorithms for computing the unique root, when one exists.

We note that Eq. (12) showed that the effect of including a linear term $\mathbf{g}_{10}'\mathbf{z}_{10}$ in the constraint was to add to $f(\lambda)$ the term $2\lambda\mathbf{g}_{10}'\mathbf{g}_{10}$ which, being linear in λ with a positive slope, has no essential effect on the geometry of Figs. 1a and 1b or on solutions to $f(\lambda)=k$. The pathological situation when no root exists in F is described next.

3.1. Zero values of s_i

In the above, repeated values of any γ_i cause no problem and neither do zero values. There is, however, one pathological situation that deserves consideration. This is when γ_1 and/or γ_p are associated with zero values of s_1 and/or s_p. Then, the first asymptotes occur at $\lambda = 1/\gamma_a$ and $1/\gamma_b$, say, bounding an enlarged region G: $\lambda = 1/\gamma_a \leq 0 \leq 1/\gamma_b$ containing F. We refer to those asymptotes that disappear as *phantom asymptotes*. The region of admissible solutions remains unchanged and $f(\lambda)$ remains strictly monotone increasing but now in the region G. There is, at most, one root λ_0 in G and if this root is also in F, it generates the desired minimum. If λ_0 is outside F, this root is a saddle point. When λ_0 does not exist, or is outside F, it may be shown that the unique minimum is given by setting/replacing λ_0 by $\lambda = 1/\gamma_1$ or $1/\gamma_p$ (see Albers et al., [1] for details). This entails setting $z_i = 0$ for all $s_i = 0$, other than z_1 or z_p, whichever is selected and whose value is to be determined. For the selected z the constraint becomes

$$f(1/\gamma) + \gamma z^2 = k, \tag{28}$$

so determining z. If λ_0 is not in F then both $f(1/\gamma_1) - k$ and $f(1/\gamma_p) - k$ must be of the same sign. When this sign is positive, γz^2 in (28) must be negative, and we choose $\gamma = \gamma_1$, so estimating z_1, else $\gamma = \gamma_p$, so estimating z_p. Thus (28) has a solution for either z_1 or z_p but not both.

An important special case is when $\mathbf{s} = 0$, so that all s_i vanish, giving $\gamma z^2 = k$. Thus, for the constraint to be satisfied, γ must have the same sign as k. We have arranged that k be positive so $\lambda = 1/\gamma_p$ with $z_p^2 = k/\gamma_p$ else $z_i = 0$ is the only possible solution. This case corresponds to the classical two-sided algebraic eigenvalue problem with minimum at the extremity of the shortest real principal axis.

These results derive from the properties of the normal equations $\mathbf{z} - \mathbf{s} = \lambda\Gamma\mathbf{z}$ which now divide into two forms:

$$\begin{array}{l}\text{(i) } z_i - s_i = \lambda\gamma_i z_i, \quad s_i \neq 0 \\ \text{(ii) } z_j = \lambda\gamma_j z_j, \quad s_j = 0.\end{array} \tag{29}$$

The first is of the kind discussed above, requiring solutions to $f(\lambda) = k$ (see below for $k = 0$). The second occurs only when $s_j = 0$ and is a simple eigenvalue expression, giving $\lambda = 1/\gamma_j$ with z_j undetermined, or $z_j = 0$ with λ undetermined. When λ is determined from (i), it follows that $z_j = 0$ is the only possible solution to (ii), but when λ is determined from (ii), solutions for z_i may be derived from (i), choosing, if possible, z_j to satisfy the constraint (28). Where no admissible z_j exists, the value of λ is not in the admissible region and must be rejected. A value of λ in the admissible region always exists that either satisfies (i) or (ii) but not both. Thus, as well as the previously discussed generalisations, (i) includes elements of conventional eigenvalue problems, with which it coincides when $\mathbf{s} = 0$.

In the above, should γ_1 (or γ_p) be a repeated root, not all of whose manifestations correspond to phantom asymptotes, then the real asymptote characteristics dominate.

3.2. The case $k = 0$

Essentially, this case is already covered by our previous development. However, there are two particular instances of $k = 0$ in (1) where it is possible, and computationally more convenient, to avoid constructing the GCF.

If $\mathbf{B}$ is p.d. and $\mathbf{b}$ is zero, the constraint $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$ is only satisfied by $\mathbf{x} = 0$ which is therefore uniquely optimal. Again, if $\mathbf{B}$ is p.s.d. and $\mathbf{b}$ is zero, the constraint $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$ is equivalent to the linear constraint $\mathbf{B}\mathbf{x} = 0$. That is, to $\mathbf{x} = \mathbf{V}\mathbf{v}$ for some $\mathbf{v}$, where the columns of $\mathbf{V}$ form a basis for the null space of $\mathbf{B}$. Thus, we have only to solve the generalised least-squares problem of minimising $(\mathbf{V}\mathbf{v} - \mathbf{t})'\mathbf{A}(\mathbf{V}\mathbf{v} - \mathbf{t})$, giving $\hat{\mathbf{v}} = (\mathbf{V}'\mathbf{A}\mathbf{V})^{-1}\mathbf{V}'\mathbf{A}'\mathbf{t}$ and, hence, $\hat{\mathbf{x}} = \mathbf{V}\hat{\mathbf{v}}$, when $\mathbf{A}$ is p.d. When $\mathbf{B}$ is indefinite our solution remains valid since, although $\mathbf{B}\mathbf{x} = 0$ implies $\mathbf{x}'\mathbf{B}\mathbf{x} = 0$, there are other vectors, not in the null space of $\mathbf{B}$, satisfying the constraint.

3.3. Algorithmic issues

Algorithms have been published for special cases of (1). Here, our objective is to outline a general algorithm that subsumes all the variants of (1) discussed above. The basic approach is to derive the GCF and solve the FCE to give the root λ of $f(\lambda) = k$, thus obtaining a minimum of (1). Various numerical difficulties can occur, due to e.g. very large/small numbers and near-singular matrices. These should be detected, and warnings given.

The root λ in (27) has to be found via numerical methods (at least when $p > 3$). Newton–Raphson or other gradient methods should be avoided because they have the potential for converging to the wrong branch of $f(\lambda)$ due to the flatness of $f(\lambda)$ for small values of λ (cf. [4]). We have found a simple bisection method to be satisfactory, usually finding a root within a second.

For the p.s.d. case (see Fig. 1), a lower bound is not directly available. Gower and Dijksterhuis[8] provide a lower bound to the admissible region, allowing the bisection method to be initiated.

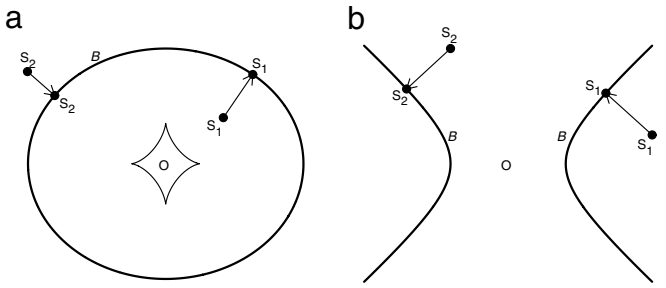

Fig. 2. In (a) B is elliptical and $\mathbf{B}$ is positive definite. In (b) B is hyperbolic and $\mathbf{B}$ is indefinite. Shortest normals are indicated for internal (s_1) and external (s_2) settings of $\mathbf{s}$. The arrows at the end of the normal indicate the constrained solutions $\mathbf{z}$. Points inside the star shaped region have four real normals to B, those outside have two.

4. Geometry

Geometrical considerations give insight into difficulties that may be encountered when minimising (10). The generality of the criterion makes it quite difficult to illustrate the full geometry in the two dimensions of a sheet of paper. We have to cope with the possibilities that $\mathbf{A}$ and $\mathbf{B}$ may be of different ranks and have different null spaces, requiring a minimum of four dimensions, as implied by the parameters $\mathbf{z}_{11}$, $\mathbf{z}_{10}$, $\mathbf{z}_{01}$, $\mathbf{z}_{00}$ of the GCF(s). Further we need to consider the effects of including the linear terms in the constraint. Thus, we can never visually represent the full generality. Nevertheless, much can be done with two dimensions and representations of three dimensions in two. Firstly we consider the geometry of the quadratic forms themselves and secondly the geometry of the FCE $f(\lambda) = k$.

4.1. Geometry of the quadratic forms

In this section we refer to the geometrical object representing the constraint by B, reserving $\mathbf{B}$ for its matrix form. If $\mathbf{B}$ is p.s.d., the quadratic surface is ellipsoidal but in general B will have elliptic, hyperbolic or, when there are linear terms in the constraint, parabolic cross-sections. Because the left-hand side of (10) is a simple squared Euclidean distance, the basic criterion is equivalent to finding the foot, S, of the shortest normal from $\mathbf{s}$ to the surface of B, thus putting (1) firmly into the class of constrained least-squares problems; any linear terms make no material difference to this interpretation. Without the condition that $\mathbf{A}$ has no negative eigenvalues, $(\mathbf{z}_1 - \mathbf{s}_1)'(\mathbf{z}_1 - \mathbf{s}_1)$ would have to be replaced by a hyperbolic distance and the least-squares rationale sacrificed.

4.2. The simplest case

The simplest case is when $\mathbf{A}$ and $\mathbf{B}$ share the same dimensions, in which case $\mathbf{B}$ has no extraneous variables and only Γ_1 is non-null in the GCF. Note that this diagonality implies that B is referred to its principal axes. This is illustrated in Fig. 2 where B is an ellipse and $\mathbf{s}$ may be any point in the same plane. The shortest normals for two settings of $\mathbf{s}$ are shown in Fig. 2a. The hyperbolic case is shown in Fig. 2b and introduces no additional problems.

There may be more than one normal, but we require only the shortest and this is unique except when $\mathbf{s}_k$ lies on a principal axis of B giving two solutions. The other normals are of interest in understanding the FCE and its solution as discussed in Section 4.3.

The star shape (the evolute of the ellipse) shown in Fig. 2 separates the regions inside B for which there are two and four real normals. For any point on the minor axis, the foot of the shortest normal is always at one end of this axis. On the major axis there are two regions: (i) between the origin and the cusp there are four normals to B and (ii) at the cusp there are three coincident normals at one end of the major axis and another normal at the other end; beyond the cusp there are only two normals (either end of the axis). These geometrical properties are associated with the "phantom asymptote" effect appearing in the solution to the FCE, discussed in Section 3. When $\mathbf{s}$ is at the origin, we have conventional eigenvalue problems, and when B is circular, any of the infinitely many points on B is a solution.

4.3. Exact solutions and inequality constraints

To understand the geometry of exact solutions it is important to distinguish points that are inside from those that are outside B. We focus on elliptical B.

One way exact solutions can occur is when the equality constraint in (1) is replaced by inequality, such as $\mathbf{x}'\mathbf{B}\mathbf{x} + 2\mathbf{b}'\mathbf{x} \leq k$. Then, if $\mathbf{s}$ is inside B the constraint is satisfied by the exact solution $\mathbf{x} = \mathbf{s}$ (see E in Fig. 3). If $\mathbf{s}$ is outside B, we require the nearest normal, as before, in which case the solution is inexact but the constraint is satisfied with equality. Similar remarks apply when $\mathbf{x}'\mathbf{B}\mathbf{x} + 2\mathbf{b}'\mathbf{x} \geq k$. It follows that inequality constraints can be handled within the same framework as equality constraints.

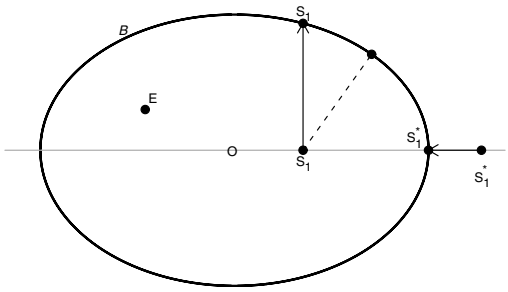

Fig. 3. The target space is one dimensional (the major axis of the ellipse). When **s** is inside there is an exact solution S that projects onto **s**. Although S itself is two dimensional, only its first coordinate is relevant. E is inside B so it gives an exact solution when the constraint is $\mathbf{x}'\mathbf{B}\mathbf{x} \leq k$.

In Section 2 we discuss several other ways in which (1) may have exact solutions. In particular, we shall be interested in when **s** is constrained to lie in a subspace of B and is inside B. We call this subspace the *target space*. In Fig. 3, B is represented as an ellipse, with its major axis representing a target space, to which **s** is confined; that is, **A** has rank one. The apparently small change to the geometry has a fundamental affect on the minimisation problem, which becomes:

$$\min(z_1 - s_1)^2 \quad \text{subject to } \gamma_1 z_1^2 + \gamma_2 z_2^2 = c^2.$$

Thus, the constraint now includes an extraneous variable (z_2) that is not part of the objective function. Indeed, γ_1 is an element of Γ_1 and γ_2 is an element of Γ_0 so now the GCF includes both diagonal matrices. Clearly, we may set $z_1 = s_1$ provided $\gamma_2 z_2^2 = c^2 - \gamma_2 s_1^2$ has a real solution for z_2; this is the condition that $\mathbf{s} = s_1$ lies inside B.

With $\mathbf{s}_1$ as shown inside B, S_1 is a point on B that projects to a point that fits **s** exactly in one dimension. In this way $\mathbf{z} = \mathbf{s}_1$. There is a second equally valid choice S_1^* at the other side of the major axis but it corresponds to the same $\mathbf{z}_1 = \mathbf{s}_1$ in the target space; z_2 is relevant only in ensuring that the constraint is satisfied. If **s** were not confined to the subspace, the shortest normal would be as indicated by the dotted line. When $\mathbf{z} = \mathbf{s}_1^*$ is outside B but in the target space, there is a unique solution at S_2 at an end of the major axis.

When B is three dimensional, but **A** remains one dimensional, then the set of points S that project into a given interior point **s** form an elliptical cross-section of the ellipsoid B, providing an infinity of exact solutions. If **s** were outside B, but remaining in the subspace, exact solutions would not exist and the approximate solution S would continue to be at one end or the other of the major axis as shown by $\mathbf{s}_1$ and S_1^* in Fig. 3. When B is three dimensional and $\mathbf{s}_2$ is confined to a target space of two of these dimensions, exact solutions remain available when **s** is inside B. When $\mathbf{s}_2$ is outside B we require the shortest normal S_2 from $\mathbf{s}_2$, to the projection of B onto the target subspace as shown in Fig. 4. This solution is not exact and, indeed, coincides with that given by the FCE, replacing B by its projection onto the target space (Appendix).

4.4. Essentially exact solutions

Consider now $\min((z_1 - s_1)^2 + (z_2 - s_2)^2)$ subject to $z_1 z_2 = c^2$, a rectangular hyperbola. This is shown in Fig. 5 but not referred to principal axes, so is not in its GCF. The GCF may be obtained by setting $\eta_1 = z_1 + z_2$ and $\eta_2 = z_1 - z_2$, so transforming the problem into $\min((\eta_1 - \frac{1}{2}(s_1 + s_2)^2) + (\eta_2 - \frac{1}{2}(s_1 - s_2)^2))$ subject to $\eta_1^2 - \eta_2^2 = 4c^2$. This is now in the GCF with the same parameters in the constraint as in the objective function so that only Γ_1 is non-null. It raises no problems and has well-defined shortest normals as indicated in Fig. 5 for various settings s_1. Note that the two-dimensional solution for the origin is well defined, up to a reflection. However, when the term $(z_2 - s_2)^2$ is excluded from the objective function, so that we seek a one-dimensional target solution, $\min((z_1 - s_1)^2)$ subject to $z_1 z_2 = c^2$, the specification is already in GCF but the diagonal matrices Γ_1 and Γ_0 both vanish and are replaced by a cross-product term corresponding to $\mathbf{D}_{11}$. This is a simple instance of where **A** and **B** are not simultaneously diagonisable. The parameter z_2 of the constraint is not in the target space, so the usual exact solutions are available (e.g. s_2 in Fig. 5) with the exception that for $\mathbf{s} = 0$, the origin, S is only defined asymptotically (termed an essentially exact solution by Albers et al. [1]). Thus, what may seem a trivial difference between two simple minimisation problems, can have a profound effect on the geometry, and hence the algebraic structure, of its solution.

Albers et al. [1] give necessary and sufficient conditions for essentially exact solutions to occur. In particular, they show that they can occur only when **B** is indefinite and $\mathbf{D}_{11}$ and/or $\mathbf{D}_{10}$ occur. In practice, such pathological solutions are probably mainly of interest in indicating that one may be attempting to fit an inappropriate model. Algorithms should trap and report on these situations.

4.5. Including linear terms

Next we consider a simple example that includes a linear term: $\min((z_1 - s_1)^2 + (z_2 - s_2)^2)$ with the constraint $z_2^2 = 4cz_1$. B is a parabola as shown in Fig. 6. We have not included a constant term but if we had this would merely shift the vertex of the parabola away from the origin O. We see that without extraneous variables the shortest normal is well defined ($\mathbf{s}_1$). If we

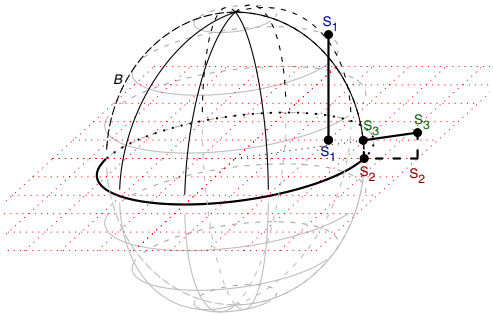

Fig. 4. B is three dimensional while the target space containing s is two dimensional (shown as the grid plane). When s is inside B there is an exact solution. When s is outside we require the shortest normal to the projection of B onto the target space. S_2 and s_2 are in a one-dimensional target space, S_3 and s_3 are in two dimensions.

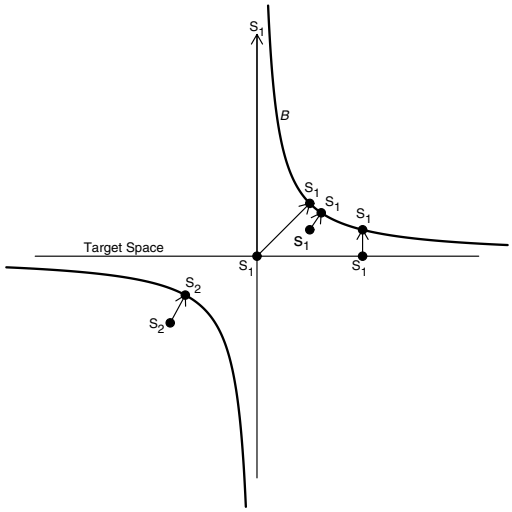

Fig. 5. A rectangular hyperbola of the form $z_1z_2 = c^2$ not referred to principal axes. Assorted two-dimensional solutions are shown together with exact solutions when the target space is restricted to the horizontal axis. The particular one-dimensional solution for $\mathbf{s} = 0$ only gives an asymptotic "exact" solution. The diagram does not correspond to the GCF for two-dimensional solutions but does for the one-dimensional solution.

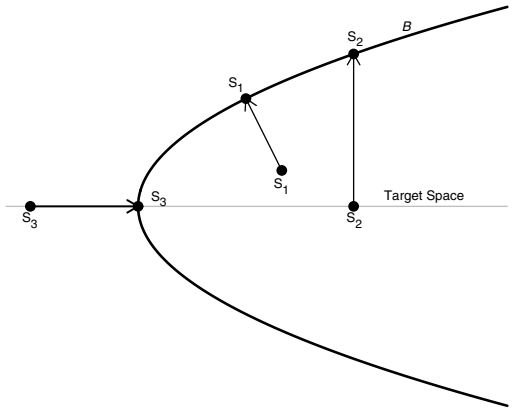

Fig. 6. B is the parabola $z_2^2 = 4cz_1$ containing a linear term. The shortest normal remains well defined (as for s_1) and when there are extraneous variables with $\mathbf{z}$ confined to the horizontal axis (as for s_2), exact solutions exist. When s is outside the parabola shortest normals remain available but when extraneous and negative only the non-exact solution at the origin O (as for s_3) is available. The linear term has had no substantive effect.

require $\min(z_1 - s_1)^2$ subject to $z_2^2 = 4cz_1$, so we have a one-dimensional target space and z_2 is extraneous, we have exact solutions when s_1 is non-negative, as shown for S_2, while if s_1 is negative we have the shortest normal at O ($z_1 = z_2 = 0$). The linear term has not raised any new problems and, as shown in Section 3, has little effect on algorithms.

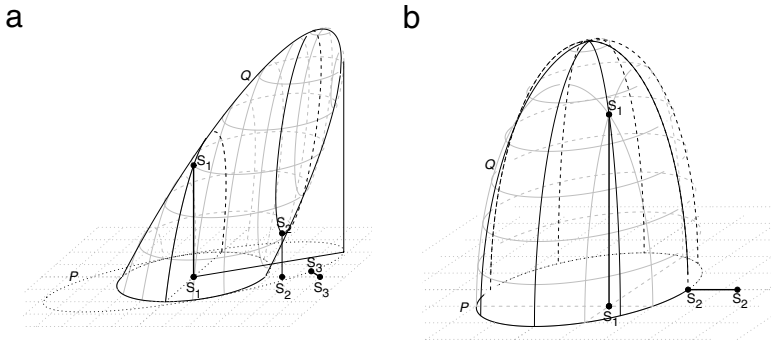

Fig. 7. The projection P of Q onto a subspace. The large dotted ellipse indicates the projection P and the smaller one the intersection of Q with a two-dimensional subspace. The algebraic form of P is given by (3) of Section 2. In (a) if the vertical axis represents an extraneous variable, points $\mathbf{s}_1$ and $\mathbf{s}_2$ indicate exact solutions when $\mathbf{s}$ lies within the intersection, though possibly outside Q itself ($\mathbf{s}_2$). The point $\mathbf{s}_3$ in the solution subspace is outside the intersection and its solution is given by the shortest normal onto P. (b) is the case when $\mathbf{C}_{12}$ vanishes, as when $\mathbf{B}$ is p.s.d. Now Q is normal to the subspace P.

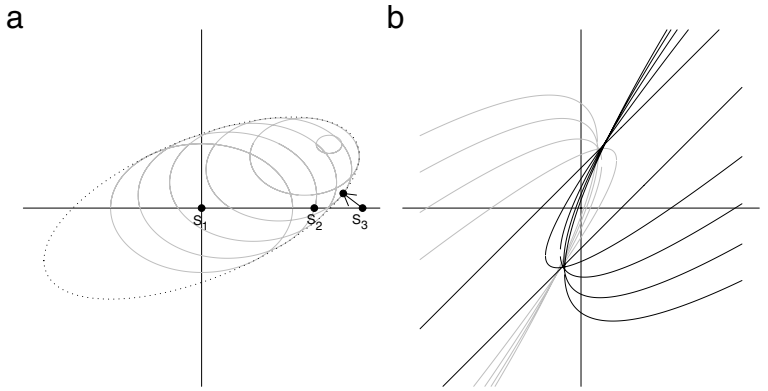

Fig. 8. (a) is a contour version of 7(a) for z_3 positive (to include z_3 negative would entail confusing overlap). (b) is a contour version when Q is hyperbolic. The darker contours are for z_3 positive and the lighter contours for z_3 negative.

4.6. General projection of a quadratic onto a subspace

The matrix $\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}$ occurs in Eq. (3) as a step in developing the GCF, and is familiar in various forms of canonical analysis. Its geometrical interpretation is that the quadratic form

$$P : \mathbf{x}_1' \left(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\right) \mathbf{x}_1 = 1$$

is the boundary of the projection of the quadratic form[1]

$$Q : \begin{pmatrix} \mathbf{x}_1' & \mathbf{x}_2' \end{pmatrix} \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = 1$$

onto the subspace spanned by $\mathbf{x}_1$; a proof is given the Appendix. Note that we are not concerned with the projection of interior points, only with their convex hull. Fig. 7a shows this for Q in three dimensions and P in two dimensions. Note that P contains, and indeed is tangential to, the intersection $\mathbf{x}_1'\mathbf{C}_{11}\mathbf{x}_1 = 1$ of Q with the subspace. This implies that when $\mathbf{s}$ lies in the same subspace as P and $\mathbf{x}_2$ corresponds to extraneous variables, then exact solutions to (1) arise not only from those $\mathbf{s}$ that are inside the intersection of Q with the subspace but also from those that lie between P and the projection of Q, as is shown in Fig. 7. When Q is p.s.d. we show in Section 2 that only diagonal values appear in the GCF and then $\mathbf{C}_{12}$ vanishes. Then, Q is normal to the subspace and the projection and intersection coincide as in Fig. 7b. Thus, the full generality occurs only when there are extraneous variables and $\mathbf{D}_{10}$ and $\mathbf{D}_{00}$ do not vanish, in which case B is not p.s.d. and must have hyperbolic cross-sections. We cannot show this in three-dimensional form but Fig. 8 gives two-dimensional contour plots. Fig. 8a is the contour version of Fig. 7a, while Fig. 8b shows contours when Q is hyperbolic. As with the rectangular hyperbola of Fig. 5, exact solutions will exist everywhere in the two-dimensional subspace, as accords with Section 2.3.

[1] Algebraically, the vector $(\mathbf{x}_1', \mathbf{x}_2')'$ projects onto $(\mathbf{x}_1', \mathbf{0}')'$, identified here with $\mathbf{x}_1$.

4.7. Summary

If we denote the space occupied by $\mathbf{s}$ as A and the constraint space by B, now using B_Q for the quadratic surface itself, the above geometry may be summarised as follows. Recalling that a proper subspace is one of strictly lower dimension, we distinguish the following four mutually exclusive and exhaustive cases:

(1) When A and B are the same space, then the solution $\mathbf{z}$ is given by the foot of the shortest normal from $\mathbf{s}$ to B_Q.
(2) When B is a proper subspace of A, then project $\mathbf{s}$ onto B to obtain $\mathbf{s} = \mathbf{s_B} + \mathbf{s_{B^\perp}}$, where $\mathbf{s_{B^\perp}}$ is the part of $\mathbf{s}$ that is orthogonal to B. Then $\mathbf{s_B}$ is in B and we may proceed as in 1 to obtain $\mathbf{z_B}$ and, finally, $\mathbf{z} = \mathbf{z_B} + \mathbf{s_{B^\perp}}$.
(3) When A is a proper subspace of B, then (a) if $\mathbf{s}$ is outside B_Q then $\mathbf{z}$ is the foot of the shortest normal from $\mathbf{s}$ onto B_Q (on a principal axis or in a principal subspace), (b) if $\mathbf{s}$ is inside B_Q then the solution is exact and $\mathbf{z}$ is the set of all points on B_Q that project into $\mathbf{s}$ (*not* the shortest normal from $\mathbf{s}$ to B_Q).
(4) When A and B intersect in a proper subspace of each of them, then project $\mathbf{s}$ onto $A \cap B$ to obtain $\mathbf{s} = \mathbf{s_{A\cap B}} + \mathbf{s_{B^\perp}}$ where $\mathbf{s_{B^\perp}}$ is again orthogonal to B. Proceed as in 3 for $\mathbf{s_{A\cap B}}$ to give $\mathbf{z} = \mathbf{z_{A\cap B}} + \mathbf{s_{B^\perp}}$.

4.8. Geometry of the FCE

Fig. 2a shows a star shape separating the region of an ellipse where there are four normals from the region where there are only two real normals. In general, the normals occur in pairs accounting for the parabolic-like branches contained in the different regions bounded by successive asymptotes of $f(\lambda)$ shown in Fig. 1. A parabolic-like branch that crosses or touches the axis in Fig. 1 gives a pair of real normals, else the paired normals are not real. Any point on the boundary of the star shaped region generates a pair of equal length normals and gives equal roots manifested by the curve $f(\lambda) = k$ touching the axis in Fig. 1. Non-parabolic shaped branches occur at either extreme region and, crucially, in the admissible region containing the origin. This behaviour of $f(\lambda)$ is shown in Fig. 9, where a transect $s_1 = s_2$ is taken through the ellipse and studied. As $s_1 = s_2$ increases, the middle "parabola" initially crosses the λ-axis, then touches it at the point corresponding to B on the star, then moves away from the axis leaving only two real roots of $f(\lambda) = 1$. On the surface of the ellipse, point D, there is an exact fit at $\lambda = 0$, the root λ being positive in the interior and negative outside the ellipse. All roots are in the admissible region. If we plotted λ against $s_1 = s_2$ the curve would be smooth. The function is well behaved but we have seen in the algebraic treatment that when $s_k = 0$ except for s_1, the behaviour of the optimal value of z_1 is more complicated. Fig. 10 shows what happens as one traverses the major axis from the origin outwards. Fig. 10a shows three versions of $f(\lambda)$, one giving a root in the admissible region, one not, and one with the root at the boundary of the admissible region. In fact for the constraint $\gamma_1 s_1^2 + \gamma_2 s_2^2 = 1$ we have that:

$$f(\lambda) = \frac{\gamma_1 s_1^2}{(1-\lambda\gamma_1)^2} = 1$$

so that $f(\lambda)$ increases linearly with s_1^2. When $\lambda \le 1/\gamma_2$, $f(\lambda) = 1$ has a root in the admissible region, i.e. when $s_1 \ge \delta = \sqrt{\gamma_1}(\frac{1}{\gamma_1} - \frac{1}{\gamma_2})$, the cusp of the star shaped region on the major axis. Fig. 10b shows how the root λ changes with s_1, being constant at $1/\gamma_2$ until the cusp δ is reached, after which it decreases linearly. However, the primary interest is in the fitted values of z_1 and z_2. These are shown in Fig. 10c where z_2 falls elliptically until the cusp, after which it becomes zero (with $z_1 = \sqrt{1/\gamma_1}$) which is at the end of the major axis. Rather counter-intuitively, the initial region for constant λ corresponds to the region of the ellipse where there are equal pairs of shortest normals from s_1 to the perimeter of the ellipse, while with variable λ after the cusp, including exterior points on the major axis, the shortest normals all end at the end of the major axis. This, and similar pathological cases are all included in the analytical treatment and the algorithm derived from it. Animations of the behaviour of the FCE are available at http://www.gmw.rug.nl/~casper/quadratic.

An example of the geometry of $k = 0$ and $\mathbf{B}$ of full rank with phantom asymptotes, is given by Albers et al. [2].

5. Discussion

At the outset we stated that our objectives were to revisit, extend and unify a rich class of optimisation problems subsumed in (1). Examples and citations given by Albers et al. [1,2] confirm that there are many applications in statistics as there probably are in other fields of science.

Previous work may be dichotomised into (i) the algebraic problems concerned with the simultaneous diagonalisation (or not) of $\mathbf{A}$ and $\mathbf{B}$ and (ii) numerical optimisation problems. The former group can be couched in quite formal mathematical language which we have tried to minimise in this article. Rao and Mitra [15] place the origins of the algebraic approach with Weierstrass and Kronecker, stating that their own approach is simpler. Yet it is formulated in terms of Hermitian matrices, generalised inverses, cogredient and contragredient transformations. This is admirably complete but is more than we have found necessary for ordinary statistical work. Thus, we are concerned only with real symmetric matrices, consider only non-singular transformations $\mathbf{T'AT}$ paired with $\mathbf{T'BT}$, and subsume generalised inverse considerations by appealing to the partition of orthogonal matrices $\mathbf{V} = (\mathbf{V}_1\ \mathbf{V}_0)$ where $\mathbf{V}_0$ represents an arbitrary set of orthogonal column vectors spanning

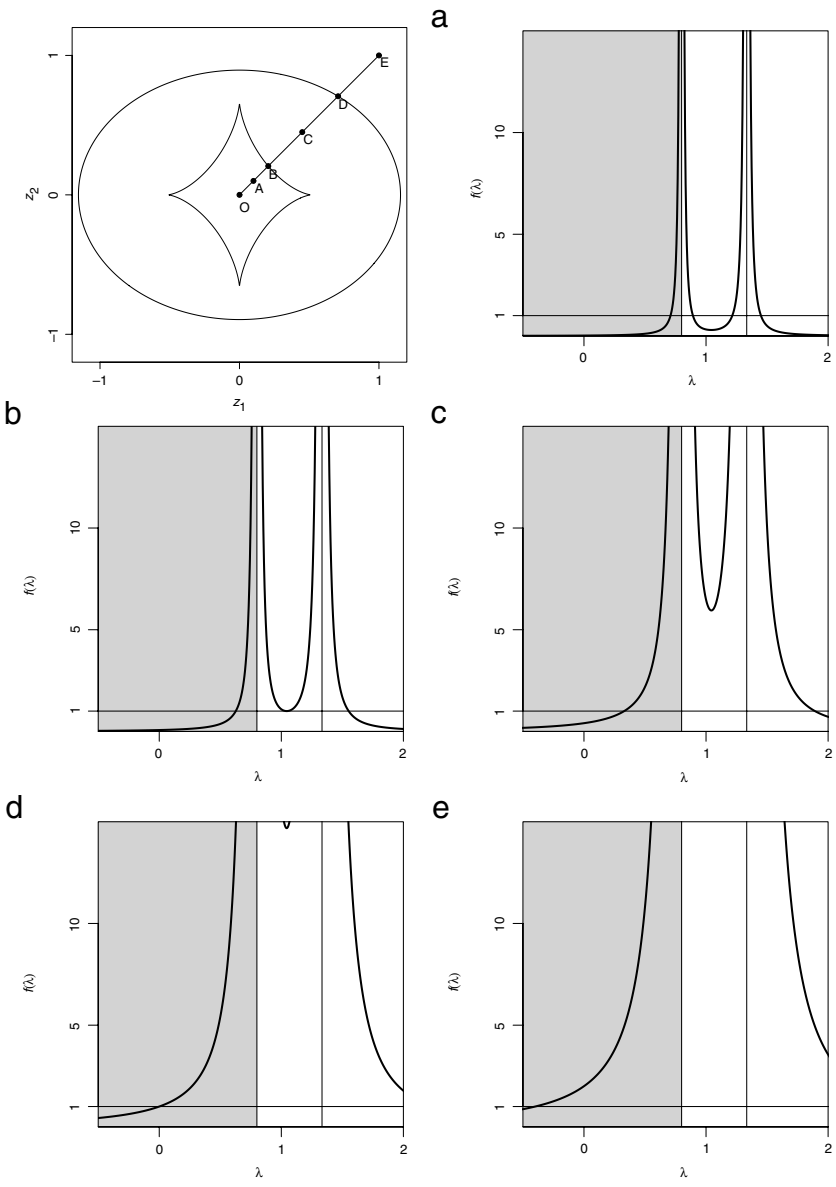

Fig. 9. Geometrical (top-left) and Lagrangian (other plots) view of the two-dimensional elliptic case, exemplified via the ellipse $\frac{3}{4}z_1^2 + \frac{5}{4}z_2^2 = 1$. The star shape in the top-left graph depicts the boundary between two and four normals.

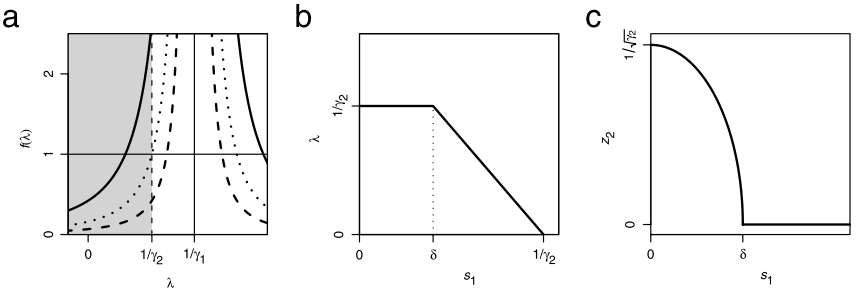

Fig. 10. (a) shows $f(\lambda)$ for different choices of $\mathbf{s} = (s_1, 0)$. The solid, dotted and dashed line correspond to $|s_1|$ greater than, equal to, or smaller than $\delta = \sqrt{\gamma_1}(1/\gamma_1 - 1/\gamma_2)$, respectively. There is a phantom asymptote at $\lambda = 1/\gamma_2$. (b) displays the relation between λ and s_1: the root of (1) is obtained at $\lambda = 1/\gamma_2$ for $s_1 < \delta$, after which it linearly decreases. (c) displays the relation between z_2 and s_1.

a null space. With these simplifications we have been able to give an explicit representation (5) for $\mathbf{T}$ and its inverse (6), which we believe to be new results.

A major concern of the statistical literature has been the optimisation of ratios of quadratic forms arising in a variety of statistical canonical variable problems. de Leeuw [10], extending work by McDonald et al. [11], discusses the optimisation of $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}$ (see Section 2.4) where $\mathbf{A}$ and $\mathbf{B}$ are p.s.d. Critchley [5] performed preliminary studies for our problem in the case where $\mathbf{A}$ is positive definite. Gower and Dijksterhuis [8] address the problem in the context of Procrustes analysis and give a preliminary algorithm.

Gander [6] studied exact solutions and their properties of the optimisation problem $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$ subject to $\|\mathbf{C}\mathbf{x}-\mathbf{d}\|^2 = 0$, that is strongly related to ours (see Section 1). More [12] studied exact solutions of minimising quadratic functions subject to ellipsoidal constraints. Minimising quadratic functions under quadratic constraints is an active topic in optimisation theory (see, for example, [16]). In general optimisation theory some variants of (1) are included in what is known as the trust region problem (see e.g. [12,3]) and ridge analysis (see e.g. [13]).

We have seen that when some elements of $\mathbf{t}$ are zero (1) has eigenvalue-like characteristics, reducing to this precise problem when $\mathbf{t} = 0$, discussed in Section 2.4. These problems may usually be expressed as two-sided eigenvalue problems, though we have seen that this is not always possible.

Notwithstanding the importance of the case $\mathbf{t} = 0$, our main concern has been with the more general problem where $\mathbf{t}$ is a given non-zero vector. Special cases have arisen in the literature but we have given a unified approach necessary for underpinning a general purpose algorithm. We believe that not only does this contribute to a better understanding of this class of problems, but also greatly helps in the formulation and solution to special cases of this class that may arise in the future. Generally there is a unique minimum to (1) but, like the algebraic eigenvalue problem, pathological solutions occur under unlikely practical circumstances that should be covered by robust algorithms. We believe that our discussion in Section 3 gives the first full treatment of solving the FCE, that takes into account all the pathological cases and the linear term in the constraint.

We have dealt with the general case where $\mathbf{A}$ is p.s.d. and $\mathbf{B}$ possibly indefinite. When $\mathbf{B}$ is indefinite, there are genuine applications when the GCF is diagonal but rarely in the more general case that includes the theoretically possible off-diagonal matrices $\mathbf{D}$. We have also included the linear term $2\mathbf{b}'\mathbf{x}$ in (1) but have seen that often reparameterisation eliminates this term. The main exception is when there is a linear term in the constraint with no quadratic counterpart, in which case the geometric form of the constraint is parabolic, which we have seen gives rise to no special problem. The other exception is when there is a linear term in the extraneous variable $\mathbf{z}_{00}$, also associated with $\mathbf{D}$. Normally $\mathbf{z}_{00}$ is immaterial but it could occur as an extraneous linear variable in the constraint, in which case it may always be chosen to give an exact fit. This draws attention to a limitation of (1) where, if the full generality is used, overparameterisation may lead to exact solutions.

Problem (1) is a vector minimisation problem. In Section 1 we showed that even an apparent multidimensional matrix generalisation involving terms such as $\min \operatorname{trace}(\mathbf{X}-\mathbf{T})'\mathbf{A}(\mathbf{X}-\mathbf{T})$ subject to $\operatorname{trace}(\mathbf{X}'\mathbf{B}\mathbf{X}) = k$ is subsumed in (1). This type of multidimensional generalisation should be distinguished from higher-dimensional solutions associated with many well-known multivariate methods based on combining solutions given by several eigenvalues. Minimising (1), as with eigenvalue problems, also has several extrema but we do not know whether these multiple solutions may be of practical interest; we suspect not because even when $\mathbf{B}$ is p.s.d. some roots of the FCE (27) may not be real. Finally, we should mention a totally different kind of generalisation of (23), involving functions of quadratic forms, with or without constraints. Thus, Kiers [9]) discusses minimising/maximising sums of ratios of quadratic forms, offering algorithmic solutions to several variants of the problem.

Acknowledgment

We thank Chris Jones for his helpful comments.

Appendix A. Projection of a quadratic onto a subspace

Theorem. *The boundary of the projection of:*

$$\mathrm{Q} : \begin{pmatrix}\mathbf{x}_1' & \mathbf{x}_2'\end{pmatrix}\begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix} = 1 \tag{A.1}$$

onto the space spanned by $\mathbf{x}_1$ *is given by*

$$\mathrm{P} : \mathbf{x}_1'\left(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\right)\mathbf{x}_1 = 1,$$

whenever $\mathbf{C}_{22}$ *is non-singular.*

Proof. The projection is defined by the points of Q that have normals in the $\mathbf{x}_1$-space. The projection of $(\mathbf{x}_1', \mathbf{x}_2')'$ at this point of Q is $\mathbf{x}_1$. The normal at $(\mathbf{x}_1', \mathbf{x}_2')'$ to Q is proportional to the vector:

$$\begin{pmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}.$$

The component of this vector that is orthogonal to the $\mathbf{x}_1$-space must vanish, i.e.

$$\mathbf{C}_{21}\mathbf{x}_1 + \mathbf{C}_{22}\mathbf{x}_2 = 0. \tag{A.2}$$

Thus, from (A.1) and (A.2):

$$\mathbf{x}_1'\mathbf{C}_{11}\mathbf{x}_1 - 2\mathbf{x}_1'\mathbf{C}_{12}\left(\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{x}_1\right) + \left(\mathbf{x}_1'\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\right)\mathbf{C}_{22}\left(\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{x}_1\right) = 1$$

i.e.

$$\mathbf{x}_1'\left(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\right)\mathbf{x}_1 = 1$$

which is P, as was to be shown. □

Remark. In the above, we have arranged that the subspace concerned is given by the first set of variables defining Q. Any other subspace could by accommodated by first orthogonally transforming it to the leading position and using the inverse transformation to return to the original parameterisation.

References

[1] C.J. Albers, F. Critchley, J.C. Gower, Explicit minimisation of a convex quadratic under a general quadratic constraint, 2009 (in preparation).
[2] C.J. Albers, F. Critchley, J.C. Gower, Applications of quadratic minimisation problems in statistics, 2008 (submitted for publication).
[3] A.R. Conn, N.I.M. Gould, P.L. Toint, Trust-Region Methods, in: MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.
[4] E.M. Cramer, On Browne's solution for oblique procrustes rotation, Psychometrika 39 (1974) 139–163.
[5] F. Critchley, On the minimisation of a positive definite quadratic form under a quadratic constraint: Analytical solution and statistical applications, Technical Report, Department of Statistics, University of Warwick, 1990.
[6] W. Gander, Least squares with a quadratic constraint, Numerische Mathematik 36 (1981) 291–307.
[7] J.C. Gower, The role of constraints in determining optimal scores, Statistics in Medicine 17 (23) (1998) 2709–2721.
[8] J.C. Gower, G.B. Dijksterhuis, Procrustes Problems, in: Oxford Statistical Science Series, vol. 30, Oxford University Press, 2004.
[9] H.A.L. Kiers, Maximization of sums of quotients of quadratic forms and some generalizations, Psychometrika 60 (1995) 221–245.
[10] J. de Leeuw, Generalized eigenvalue problems with positive semidefinite matrices, Psychometrika 47 (1982) 87–93.
[11] R.P. McDonald, Y. Torii, S. Nishisato, Some results on proper eigenvalues and eigenvectors with applications to scaling, Psychometrika 44 (1979) 211–227.
[12] J.J. Moré, Generalizations of the trust region problem, Optimization Methods and Software II (189–209) (1993).
[13] R.H. Myers, D.C. Montgomery, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 2nd edition, in: Wiley Series in Probability and Statistics, John Wiley & Sons, New York, 2002.
[14] R.W. Newcomb, On the simultaneous diagonalization of two semi-definite matrices, Quarterly Journal of Applied Mathematics 19 (1961) 144–146.
[15] C.R. Rao, S.K. Mitra, Generalized Inverse of Matrices and its Applications, John Wiley & Sons, New York, 1971.
[16] H. Tuy, N.T. Hoai-Phuong, A robust algorithm for quadratic optimization under quadratic constraints, Journal of Global Optimization 37 (2007) 557–569.